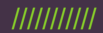




PHUSE SDE Beerse 2019

PORTIN – Next Level of Integrated Databases



September 11, 2019

Joerg Guettner





Agenda

// Introduction

// **Portin**

// What is Portin?

// How to upload data to Portin

// Web interface and API

// Conclusion



Introduction



Introduction

Integrated Databases at Bayer

- Dedicated department for more than 30 years
- Focus on Phase II to IV interventional clinical trials
- For regulatory purposes
(Integrated Summary of Safety,
Integrated Summary of Efficacy,
Health Authority requests),
but also publications
- Using SAS to create integrated databases on compound level





Introduction

How to find and analyze biomarkers not stored in the clinical database of a study?

The problem:

- Not all biomarkers are stored in the clinical database
- There are bio samples still available after study closure for some time and therefore the possibility to measure new biomarkers

How can these additional biomarkers be found and analyzed in conjunction with the clinical database?





Portal For Translational Data Integration

Data

3 Data Sources

Clinical parameters, including the age, weight and height of a patient.

Study parameters, such as the study drug, dosing regimen, targeted disease and number of subjects.



Clinical study
database (TOSCA)

Molecular biomarker data, including: mutations, gene and protein expression profiling.



Biomarker Original
Data Archive (BMDA)

Sample attributes, including the tissue (blood/tumor), expiry date and availability.



Long-term storage
Sample data (eSample)

PORTIN – source data types

Molecular Biomarker Data



DNA

digital PCR (Beaming)
DNA targeted sequencing (NGS)
mixed methods of DNA assay

RNA / miRNA

microarray
PCR-quantitative (qPCR)

Protein

immunohistochemistry
multiplex immunoassay/
ELISA

Small Molecules

liquid chromatography
mass spectrometry

Cells

situ hybridization
flow cytometry

Sample Data



Sample data

Vial Status	Visit
substance	sampld_date / time
sample_label	sample_type
study_no	expiry date
screening_no	sample_caretaker

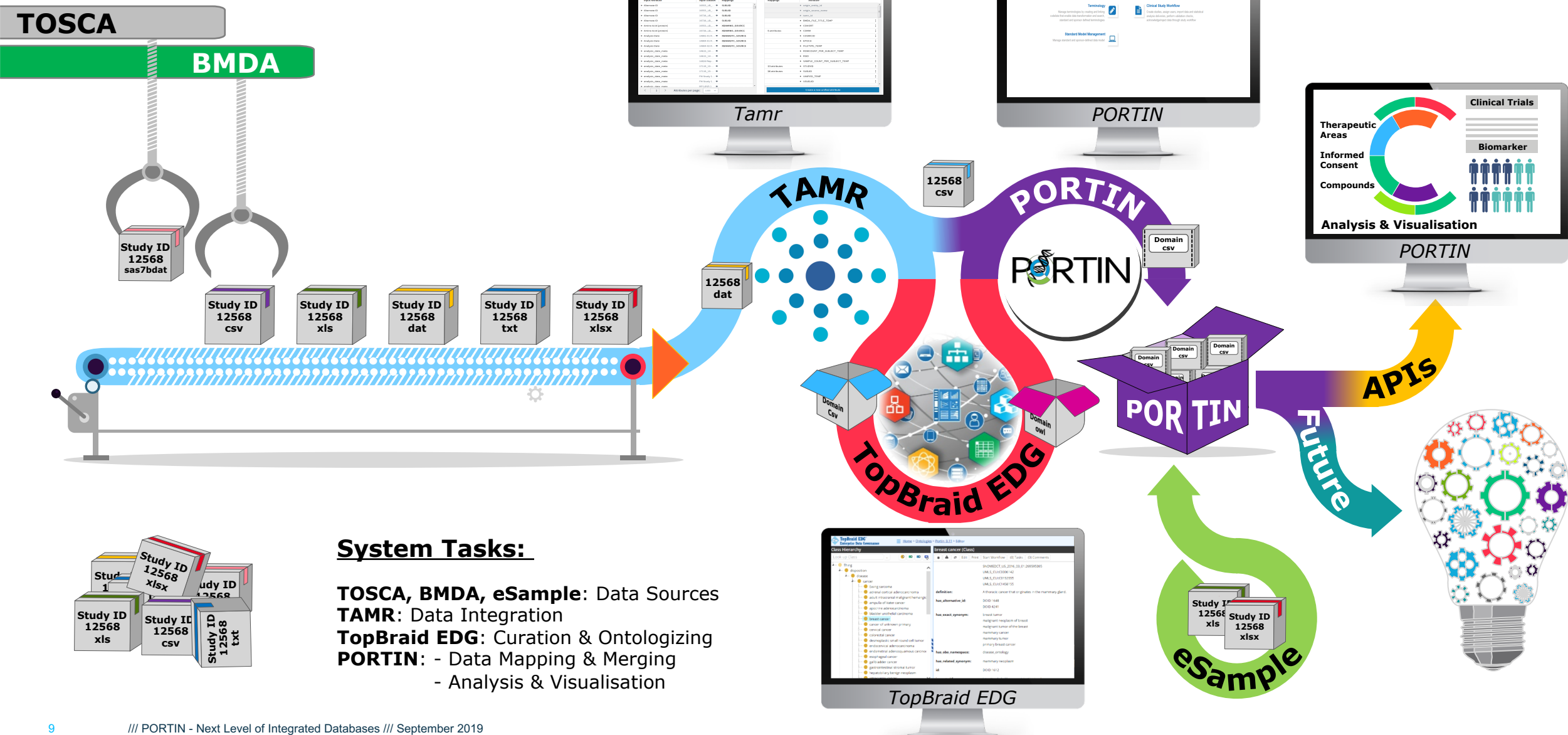
Clinical data



Domains

AE Adverse Events	EG Electrocardiogram	SU Substance Use	XB Biomarker Investigations
CE Clinical Events	EX Exposure	SV Subject Visits	XI Invalidity Reason
CM Concomitant Meds	FA Findings About Events or Interventions	TI Trial Inclusion / Exclusion Criteria	XL Cardio-Pulmonary Function Assessment
CO Comments	LB Laboratory Data	TS Trial Summary	XP Sponsor Procedures
DM Demographics	MH Medical History	TV Trial Visits	
DS Disposition	PC Pharmacokinetic Concentration	VS Vital Signs	ADTTE Time-To-Event Analysis Dataset
DV Protocol Deviations	QS Questionnaires	XA Adjunct Therapy	

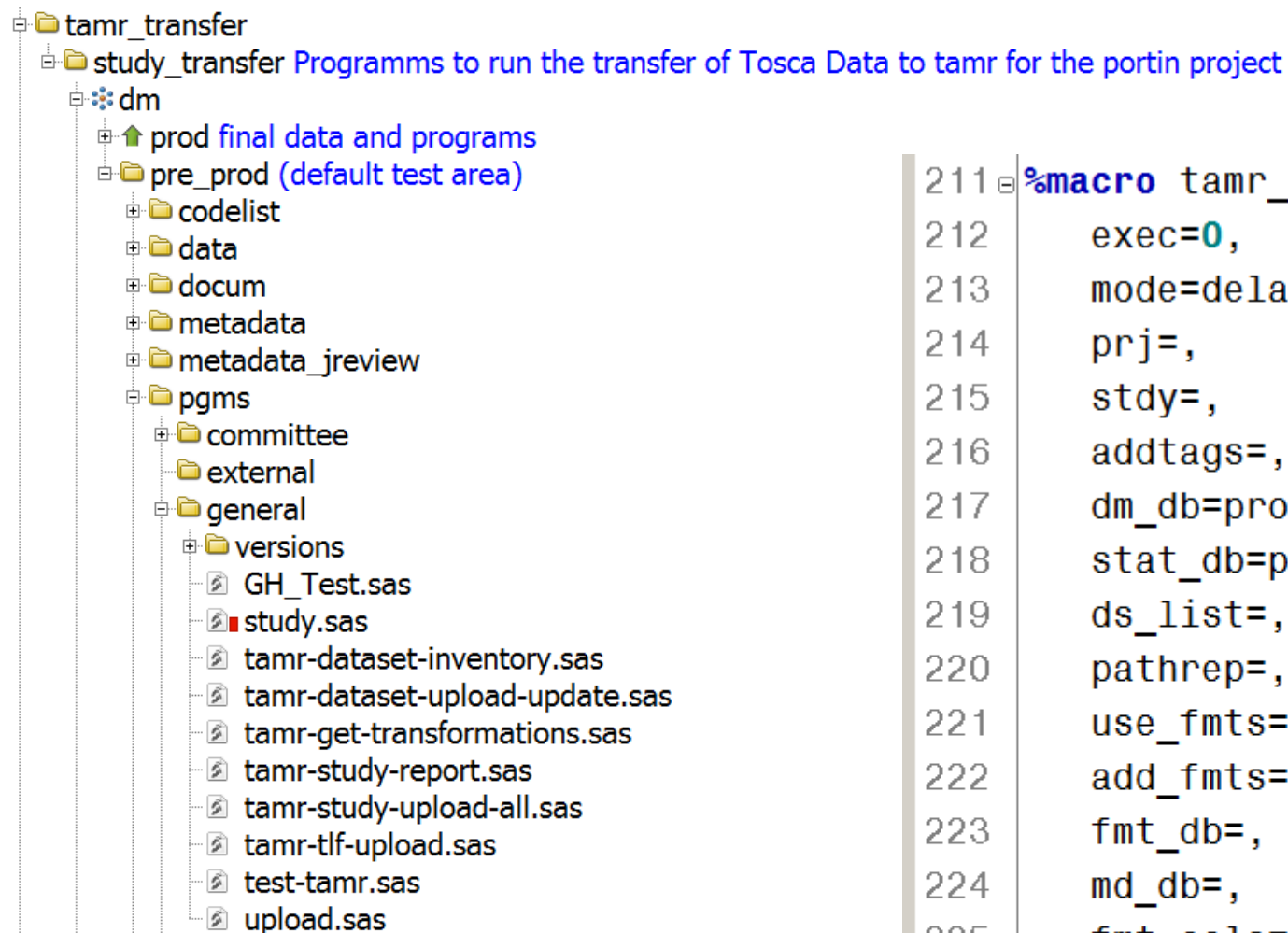
PORTIN Data Upload Process





Portin Data Upload from Tosca

Data pushed via SAS

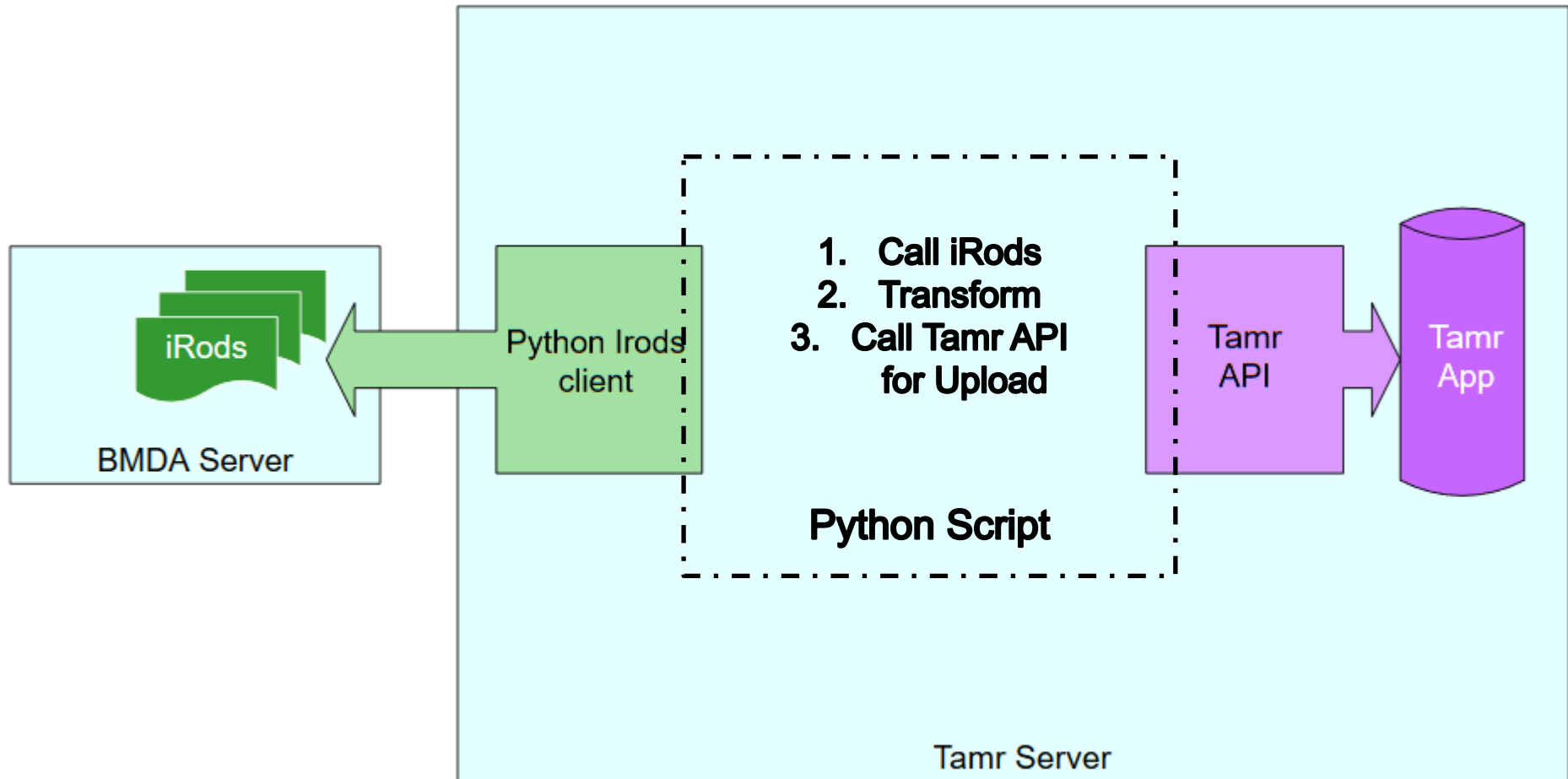


```
211 %macro tamr_add_study_gms(  
212     exec=0,  
213     mode=deladd,  
214     prj=,  
215     stdy=,  
216     addtags=,  
217     dm_db=prod,  
218     stat_db=prod,  
219     ds_list=,  
220     pathrep=,  
221     use_fmts=1,  
222     add_fmts=0,  
223     fmt_db=,  
224     md_db=,  
225     fmt_cols=,
```



PORTIN Data Upload from BMDA

Data pulled via Python script





Tamr

dna sdm Datasets Schema Mapping Unified Dataset

All Datasets JG

Sci dna sdm Datasets Schema Mapping Unified Dataset

DNA-SDM (1,000 records, 97 attributes)

This is a sample of 5 source datasets. Choose which datasets to preview or view all records.

origin_sour...	origin_entit...	tamr_id	BMDA_FILE...	COHORT	COMM	COSMICID	EPOCH	FILETYPE_T...	ORRES_temp	PAN
12871_DNA.csv	4846	221667086411...	12871_DNA.csv	null	null	COSM532	null	null		
12871_DNA.csv	633	-54071356907...	12871_DNA.csv	null	null	COSM760	null	null		
12871_DNA.csv	3205	638107974801...	12871_DNA.csv	null	null	COSM764	null	null		
12871_DNA.csv	5414	758181731056...	12871_DNA.csv	null	null	COSM516	null	null		
12871_DNA.csv	4775	-29413595119...	12871_DNA.csv	null	null	COSM516	null	null		
12871_DNA.csv	2427	527387634242...	12871_DNA.csv	null	null	COSM517	null	null		
12871_DNA.csv	4257	-37911567240...	12871_DNA.csv	null	null	COSM522	null	null		
12871_DNA.csv	2792	295174195953...	12871_DNA.csv	null	null	COSM763	null	null		
12871_DNA.csv	2350	447300754619...	12871_DNA.csv	null	null	COSM522	null	null		
12871_DNA.csv	1192	-56236065788...	12871_DNA.csv	null	null	COSM774	null	null		
12871_DNA.csv	4991	662810090834...	12871_DNA.csv	null	null	COSM775	null	null		
12871_DNA.csv	3778	-56382536942...	12871_DNA.csv	null	null	COSM518	null	null		
12871_DNA.csv	3314	872422036399...	12871_DNA.csv	null	null	COSM775	null	null		

Update Unified Dataset

Create attribute

Hide transformations

Preview all

Cancel changes

TRANSFORMATIONS

TRANSFORMATIONS FOR INPUT DATASETS

Preview without any transformations.

= <> Script

```
1 -- Cleanup gene names  
2 load from different f  
3 samples: select *, 1  
4 use  
5 "16553_LB_201512:  
6 eted Genes_toload:  
7 genes: select *,  
8 array(replace_all(ge  
9 '['^0-9a-zA-Z;,+',"')  
10 as temp;  
11 use samples;  
12 samples_genes: ful  
13 with genes on temp  
14 genes.temp;  
15 select *,  
16 genes.Gene as XBGE
```

TRANSFORMATIONS ON UNIFIED DATASET

= <> Script

```
1 -- fill these fields dow  
2 rows by XBEXTID  
3 window max(get(XBI  
4 as XBREFID,
```

TopBraid EDG for Ontologizing and Data Curation

Tamr

Standardization process and rearrangement of columns

data file

Field 1	Field 2	Ontology ID	Field 4
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx

Generating a reduced data file

Field 2	Ontology ID
xxxxxxxxxxxx	
xxxxxxxxxxxx	
xxxxxxxxxxxx	
xxxxxxxxxxxx	
xxxxxxxxxxxx	
xxxxxxxxxxxx	
xxxxxxxxxxxx	
xxxxxxxxxxxx	
xxxxxxxxxxxx	

Domain	Look up Variable	Ontology ID Variable	External ontologies used	Parent IRI	Ontology class name
DM	SEX	SEX_ONTID	1. PATO	http://purl.obolibrary.org/obo/PATO_0000047	Thing - quality - Sex
DM	RACE	RACE_ONTID	1. NCIT	http://purl.obolibrary.org/obo/NCIT_C17049	Thing - Quality - Race
DM	ETHNIC	ETHNIC_ONTID	1. NCIT	http://purl.obolibrary.org/obo/NCIT_C16564	Thing - Quality -Ethnic Group

Upload of a two column data file

Mapping process

Field 1	Field 2	Ontology ID	Field 4
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx		xxxxxxxxxxxx

Field 1	Field 2	Ontology ID	Field 4
xxxxxxxxxxxx	xxxxxxxxxxxx	100000	xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx	100001	xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx	100002	xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx	100003	xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx	100004	xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx	100005	xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx	100006	xxxxxxxxxxxx
xxxxxxxxxxxx	xxxxxxxxxxxx	100007	xxxxxxxxxxxx

PORTIN Ontology OWL file



Portin

The Web App

Browser address bar: <https://portin.intranet.cnb/SDIS/Service.svc/SDIS/0/90?CrpID=0>

Navigation: Meistbesucht | Getting Started | Latest Headlines | Aktuell - FAZ.NET | Tamr

Cohort and Biosample Selection | Support | I Wish... | Guettner, Joerg | 04-SEP-2019 07:55:53 (UTC +02:00) | Bayer

Buttons: Saved Search Criteria | Search Criteria | Show Search Results | Data Viewer

Results Cluster: Apply Filter | Clear Filter | Help

- 1 Study Drug
- 3 Study Objective(s)
- 1 Trial Phase
- 1 Project ID

A1 AND B1 AND C1 | Save Criteria | No. of Results: Studies-1; Subjects-1

Study View

Cross Study Viewer | Open Documents | IGO | Type to filter the studies | Export

☐	Study ID	Project ID	Study Title	Study Drug	Disease	Subjects Matched	Subjects with Biological Samples Available
☐	19092	194612	Single dose, open-	Finerenone		1 of 20	1 of 1

Total rows: 1



R API for PORTIN

R package for connecting R with PORTIN backend

The screenshot displays the RStudio environment. The top pane shows a table with 7 columns: STUDYID, UNIQUEID, TrialPhase, CompoundName, Indication, and StudyTitle. The bottom pane shows a terminal window with R code and its output.

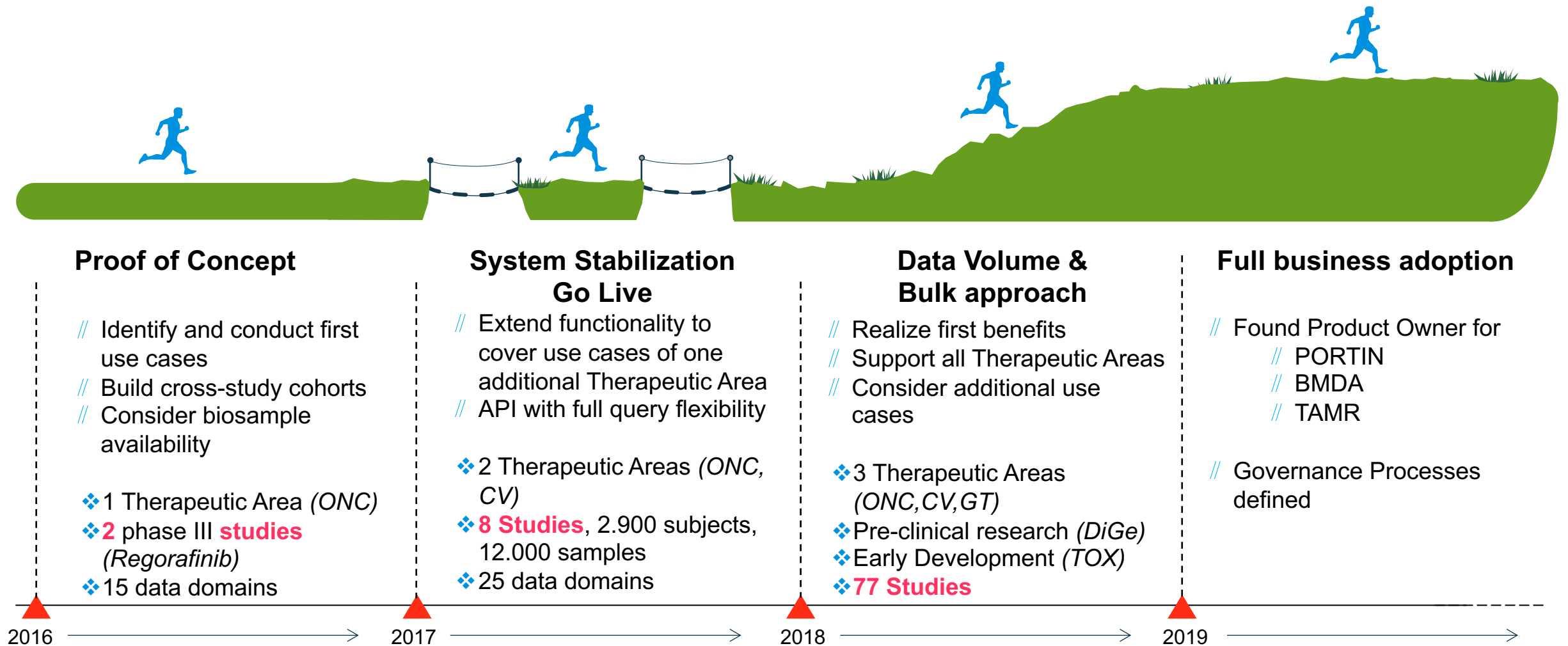
	STUDYID	UNIQUEID	TrialPhase	CompoundName	Indication	StudyTitle
1	12876	2088	Phase I Trial	Copanlisib	Advanced cancer	Phase Ib trial of the combination
2	16553	2089	Phase II Trial	Refametinib	KRAS OR NRAS MUTANT UNRESECTABLE OR METASTATIC ...	A PROSPECTIVE, SINGLE-ARM, M
3	9900012	2090			Chronic Kidney Disease	German Chronic Kidney Disease
4	14905	2091	Phase I/II Trial	Refametinib,Gemcitabine	Determine the maximum tolerated dose and the efficacy ...	A multi-center Phase I / II study
5	15734	2096	Phase I Trial	PI3Ka/Å Balanced Inhibitor	Study Part A: Advanced solid tumors, Study Part B: Locally...	An open-label Phase I dose-esca
6	14723	2098	Phase I Trial	Vilaprisan	Treatment of uterine fibroids	A multi-center, randomized, dou

```
V:/test2/
> library("xbiomAPI")
> MachineName <- "https://portin.intranet.cnb"
> Xbiom.Login(MachineName, u, pw)
[1] "successfully logged on"
> s = xbiom.ListofStudies()
[1] "Fetching next batch, Total number of batches fetched : 1"
[1] "PROCESSED ALL THE BATCH"
> view(s)
> source("helpfuns.r")
> u<-"antigoni.elefsinioti@bayer.com"
> pw=Sys.getenv("password")
>
> library("xbiomAPI")
> MachineName <- "https://portin.intranet.cnb"
> Xbiom.Login(MachineName, u, pw)
[1] "successfully logged on"
> s = xbiom.ListofStudies()
[1] "Fetching next batch, Total number of batches fetched : 1"
[1] "PROCESSED ALL THE BATCH"
> view(s)
```




PORTIN Journey

An idea becomes part of the business organization



Conclusion

- PORTIN provides harmonized, exploratory and selected clinical data to enable data insights
- Portin is a one-stop-shop where you can easily explore data from Bayer-sponsored legacy clinical trials for the benefit of your scientific hypothesis, research project or upcoming new trials



Thank you!
Questions?



joerg.guettner@bayer.com

